CLASSIFICATION OF NDE WAVEFORMS WITH AUTOREGRESSIVE MODELS

R.B. Melton

Pacific Northwest Laboratory
Richland, WA 99352

ABSTRACT

This paper describes a new approach for classifying NDE waveforms. Using this approach a set of matched filters is constructed, one for each category of waveform, based on parameters from autoregressive models. The method offers advantages in terms of hardware implementation over conventional pattern recognition approaches. Feasibility is shown using computer generated data. Results are then presented for real data from acoustic emission experiments.

INTRODUCTION

In recent years signal processing and pattern recognition techniques have been successfully applied to the problem of classifying several types of nondestructive evaluation (NDE) data, including eddy current (Doctor and Harrington, 1980), ultrasonic (Busse et al., 1981), and acoustic emission (AE) waveforms (Hutton et al., 1981, Harrington and Doctor, 1980, Melton et al., 1982). The techniques that have been developed, however, are computationally complex to the point of prohibiting real-time processing of NDT data for on-line monitoring/inspection applications. This is mainly due to the fact that the features that have been found to be important in the classification process have been calculated via transforms, such as the fast Fourier transform (FFT). For ultrasonic and acoustic emission waveforms the situation is made even worse by the high bandwidth of the data.

In this paper a new approach is described for classifying NDE waveforms. Using this approach, a set of matched filters are constructed, one for each category of waveform, based on parameters of autoregressive models of the data. This approach has advantages in

terms of real-time processing because the matched filters can be implemented as recursive digital filters. This means that the waveform data may be processed serially.

The first part of the paper presents background on autoregressive models and shows how such a model may be viewed as a recursive filter. The second section describes an experiment with computer generated data that establishes the feasibility of the approach and relates it to conventional pattern recognition. The third section describes the application of the approach to acoustic emission data. The final section discusses the results and some of the research issues that surround this approach.

BACKGROUND

Over the past 15 years new approaches to spectral analysis have been developed as an alternative to approaches using Fourier transforms. These new approaches, all of which have been shown to be equivalent (van den Bos, 1971), are called autoregressive (AR) modeling, maximum entropy spectral analysis (MESA), and linear prediction (LP). They have been developed respectively by statisticians, physical scientists, and engineers for applications within their particular areas of endeavor.

The essence of these techniques is that they allow high resolution spectral estimates to be computed from limited amounts of data by extrapolating (predicting) data values between the measured points defined by data sampling. This increases their spectral resolution, making it much greater than that of the Fourier transform techniques for any given data set (Burg, 1967, van den Bos, 1971).

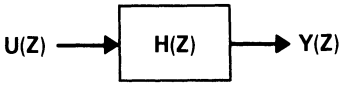
Mathematically the model is viewed in the Z-transform domain as an all-pole filter, H(z), excited by a white Gaussian process, U(z) (Makhoul, 1976). The transfer function for this relationship is written as:

$$Y(z) - U(z)H(z) \tag{1}$$

Figure 1 represents this graphically. Here U(z), H(z), and Y(z) are the complex Z-transforms of the corresponding sequences u_n, h_n, and y_n. In other words

$$U(z) = \sum_{n=-\infty}^{+\infty} (u_n Z^{-n}) \quad ; \tag{2}$$

if we let $Z=e^{jw}$, then we have the more familiar discrete Fourier transform.



$$Y(Z) = U(Z)\,H(Z)$$

Fig. 1. Transfer function relationship for autoregressive model.

Since H(z) is required to be an all-pole filter in our model, we can rewrite it as

$$H(z) = \frac{G}{A(z)} \quad , \tag{3}$$

where A(z) is a polynomial in z whose roots are the poles of the model and G is a constant gain term. Thus we may write:

$$Y(z) = \frac{GU(z)}{A(z)} \quad , \tag{4}$$

rearranging (4) we have

$$Y(z)A(z) = GU(z) \quad . \tag{5}$$

If we write out A(z) as a polynomial, we have

$$A(z) = 1 - \sum_{k=1}^{+\infty} (a_k Z^{-k}) \quad . \tag{6}$$

Substituting (6) into (5) and rearranging, we have

$$Y(z) = GU(z) + \sum_{k=1}^{+\infty} (a_k Z^{-k})Y(z) \quad . \tag{7}$$

If we now take the inverse Z-transform of both sides of (7), we have

$$y_n = Gu_n + \sum_{k=1}^{+\infty} (a_k y_{n-k}) \quad . \tag{8}$$

Thus for this modeling approach the current value of y_n is seen to depend only on the past values of y_n and a sample, u_n,

from a white noise process. In practice the model is limited to a finite number of poles and takes the form

$$y_n = Gu_n + \sum_{k=1}^{p} (a_k y_{n-k}) \qquad (9)$$

for a p-th order model.

This model is equivalent in form to a p-th order recursive filter with input u and output y. We have, then, a mathematical model of the data, y, that consists of a linear filter that, when excited by white noise, has the same spectral properties as the data. We can think, therefore, of this filter as being spectrally matched to the data.

To compute the filter coefficients Eq. (9) is thought of as a regression equation with y_n the independent variable. The dependent variables are represented by the y_{n-k}, the regression coefficients by the a_k, and the residual error by the Gu_n. Thus the name autoregressive modeling, since the data are regressed with themselves. Gu_n may also be thought of as a prediction error, thus the term linear prediction. For more details and an explanation of the term maximum entropy, the reader is referred to Childers (1978).

This leads us to the main idea in this paper, which is the use of such spectrally matched recursive filters to classify waveforms from different sources. The next section of the paper considers the use of these filters to classify computer generated data. The results of such a classification are compared to a conventional pattern recognition approach using the filter coefficients, that is the a_k, as pattern recognition features.

FEASIBILITY OF APPROACH

This section presents the results of a computer experiment designed to test the hypothesis that recursive filters whose coefficients are the average autoregressive parameters for two categories of waveforms can be used to discriminate waveforms from each category by examining the prediction error of the two filters when each is applied to an unknown waveform.

To test this hypothesis a set of test data was generated for two different autoregressive processes. Fifty waveforms were generated for each process using the following equations:

Process 1: $x1_n = 0.75*x1_{n-1} - 0.5*x1_{n-2} + w_n$

Process 2: $x2_n = 0.5*x2_{n-1} - 0.25*x2_{n-2} + w_n$

where w_n is one sample from a random number generator.

The next step in testing the hypothesis was to compute a second-order autoregressive model for each waveform. This was done using the FTMXL subroutine in the IMSL (1982) subroutine package. This subroutine implements the Box and Jenkins (1976) iterative algorithm for computing AR parameters. Figure 2 shows a cluster plot of these parameters. The vertical axis is AR parameter 1, while the horizontal axis is AR parameter 2. Note that the parameters form two distinct clusters corresponding to the two sets of test waveforms. The centers of the two clusters (computed by taking the average of each parameter within each category) are:

	AR Parameter 1	AR Parameter 2
Process 1:	0.73469	-0.49089
Process 2:	0.49413	-0.25951

Using these average AR coefficients as parameters for two recursive filters, each waveform in the test was categorized based on the difference between the prediction error of the category-two filter and the category-one filter. This is represented diagramatically in Fig. 3.

The results for classifying the waveforms with recursive filters were excellent. Figure 4 is the same as Fig. 4 but with the

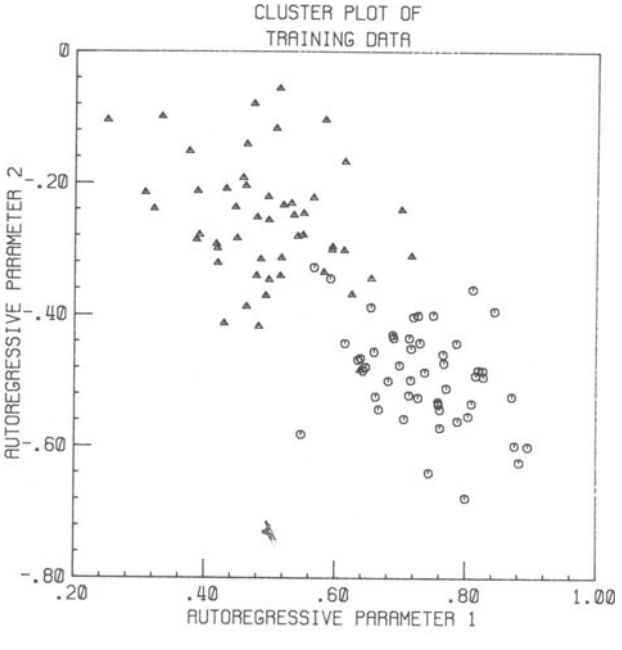


Fig. 2. Cluster plot of training data.

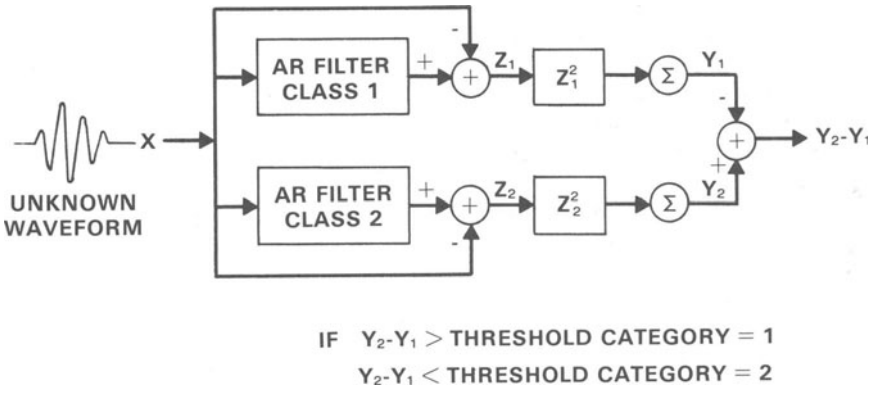


Fig. 3. Block diagram of waveform classification procedure.

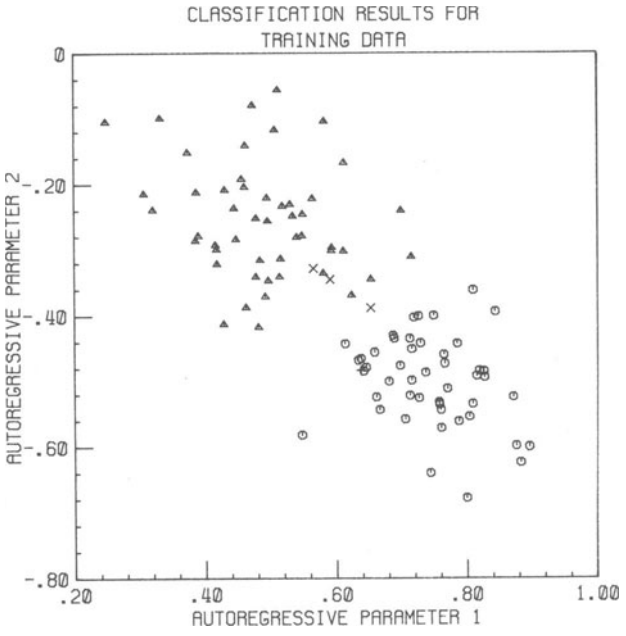


Fig. 4. Classification results for training data. X and + represent misclassified points.

misclassified waveforms identified. Note that one category-two waveform and three category-one waveforms were misclassified. This is a 96% performance rate. An independent set of 100 waveforms was generated using the same procedure outlined above for the first set but with an independent set of w(n). These data were classified using the filters derived from the first set of data. Figure 5 is a cluster plot showing the results for the independent test data. Again, the classification accuracy was 96%.

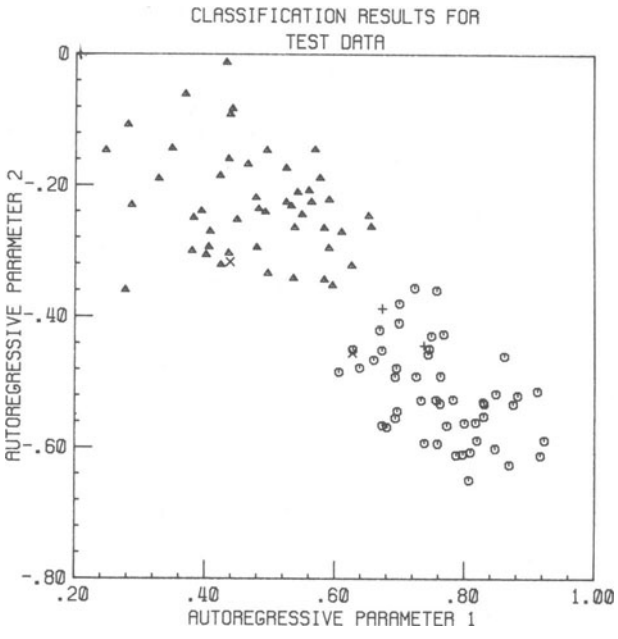


Fig. 5. Classification results for test data. X and + represent misclassified points.

APPLICATION TO ACOUSTIC EMISSION DATA

After its feasibility was established, the approach was applied to acoustic emission data to test its performance on real data. The AE data were generated experimentally using a simulated joint specimen of 7075-T651 aluminum. The specimen was configured such that it generated AE from bolt-hole fretting and crack growth. Further details have been reported earlier in these proceedings (Hutton and Lemon, 1982, Melton et al., 1982). The data were recorded using wideband piezoelectric transducers and a Biomation 1010 transient recorder with a sample rate of 5 MHz and a buffer size of 4096 points.

For this analysis 467 AE waveforms were considered, 230 fretting waveforms and 237 crack AE waveforms. For each waveform a thirtieth order autoregressive model was computed using the algorithm presented by Andersen (1974). This algorithm stems from maximum entropy method of spectrum analysis. For each class of waveforms (crack AE or fretting), the autoregressive parameters were stored for use in conventional pattern recognition and averaged to provide the coefficients for the matched filters.

The results of conventional pattern recognition were excellent. Using the autoregressive parameters as features, a multivariate linear decision rule was calculated using the ARTHUR pattern recognition

package (Duewer, et al.). In doing this, 80% of the data were used as a training set and the remaining 20% as a test of the decision rule. The results of this analysis were that for the training set, 98% of the fretting and 100% of the crack AE were correctly classified. For the test set, 100% of the fretting and 98% of the crack AE were correctly classified.

Following the conventional pattern recognition analysis, the average autoregressive parameters from a 150-waveform subset of the data was used as the coefficients for the matched filters. These filters were then used in the manner described above to classify each of the 467 waveforms, including the 150 in the "training" set. The results of this classification were that 84% of the crack AE was correctly classified and 77% of the fretting.

DISCUSSION

This paper has presented a new approach for classifying NDE waveforms. The approach consists of computing autoregressive (AR) models for each waveform in a training set of data. For each category of data the average AR parameters are computed and used as coefficients of recursive filters. Unknown waveforms are classified by running them through the filter for each category. The mean square prediction error is computed and summed for each filter. The waveform is classified as belonging to the category whose filter had the lowest prediction error.

On computer generated data the procedure was shown to compare extremely well with conventional pattern recognition using the AR parameters as features. Both achieved an accuracy of approximately 96% on both the training and test data. Based on the promise of these results, the technique was tried on real data from an acoustic emission experiment.

On the real data the procedure did not perform as well. Conventional pattern recognition using the AR coefficients as features had a classification accuracy of about 98%. The AR modeling procedure, however, only had a classification accuracy of about 80%. This points to the need for further study of this approach.

Some of the questions that need to be answered are common to this approach and modern spectrum analysis in general. The order of the AR model, that is, the number of coefficients computed, certainly affects the accuracy of the procedure. In this work the order was chosen by qualitataively comparing plots of the power spectra computed via fast Fourier transforms (FFT's) and AR models. There are more rigorous procedures, such as the Akaike Information Criteria (AIC), for determining the best order model for a given set of data (Akaike, 1974, Tong, 1975 and 1977). These procedures need to be applied to the acoustic emission data to see if the results improve.

For the computer generated test and training data, the order was known precisely since it was under our control.

There are other closely related modeling techniques that should be tried, one of which is autoregressive-moving average (ARMA) modeling (Box and Jenkins, 1976). This type of model assumes a pole-zero model instead of an all-pole model. The all-pole model can, in theory, approximate zeros (Chen, 1982), but an extremely high-order model may be required. For a waveform with a sharp low-frequency cutoff, due perhaps to transducer characteristics, the ARMA model might be more appropriate than the AR model since it could account directly for the low-frequency zeros. In the future, pattern recognition approaches based on modern spectral analysis techniques should outperform more general techniques since the physical characteristics of the system under study can be mathematically incorporated in the analysis approach.

ACKNOWLEDGEMENT

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC06-76RLO-1830.

REFERENCES

1. Akaike, H., 1974, A new look at the statistical model identification, IEEE Trans. Autom. Control, AC-19:716-723.
2. Andersen, N., 1974, On the calculation of filter coefficients for maximum entropy spectral analysis, Geophysics, 39:69-72.
3. Box. G.E.P. and Jenkins, G.M., 1976, "Time Series Analysis: Forecasting and Control," Holden-Day, San Francisco.
4. Busse, L.J., Selby, G.P., Doctor, S.R. and Becker, F.L., 1981, Natural versus machined defects: differences in ultrasonic response, Presented at 13th Symposium on NDE, San Antonio, 1981.
5. Chen, C.H., 1982, Applications to geophysics: Part I, in CRC Press, Boca Raton, Florida.
6. Childers, D.G., ed., 1978, "Modern Spectrum Analysis," IEEE Press, New York.
7. Doctor, P.G. and Harrington, T.P., 1980, Analysis of eddy current data using pattern recognition, in the Proceedings of the Fifth International Conference on Pattern Recognition, IEEE Computer Society, 137-139.
8. Duewer, D.L., Koskinen, J.R. and Kowalski, B.R., ARTHUR, availablve from B.R. Kowalski, Laboratory for Chemometrics, Dept. of Chemistry, BG-10, University of Washington, Seattle, WA.
9. Harrington, T.P. and Doctor, P.G., 1980, Acoustic emission analysis using pattern recognition, in the Proceedings of the Fifth International Conference on Pattern Recognition, IEEE Computer Society, 1204-1207.

10. Hutton, P.H., Lemon, D.K., Melton, R.B. and Doctor, P.G., 1981, Develop in-flight acoustic emission monitoring of aircraft to detect fatigue crack growth, Review of Progress in Quantitative NDE, Vol. 1, ed. D.O. Thompson and D.E. Chimenti, Plenum Pub. Corp., New Yrok, 1982.
11. Hutton, P.H. and Lemon, D.K., 1982, Development of acoustic emission methods for in-flight monitoring of aircraft structures, these proceedings.
12. IMSL Library, Edition 9, 1982, IMSL, Houston, TX.
13. Makhoul, J., 1976, Linear prediction: A tutorial review, Proc. IEEE, 63:561-580.
14. Melton, R.B., Doctor, P.G. and Daly, D.S., 1982, Pattern recognition analysis of acoustic emission waveforms from simulated joint specimens, these proceedings.
15. Tong, H., 1975, Autoregressive model fitting with noisy data by Akike's information criterion, IEEE Trans. Inform. Theory, IT-21: 476-480.
16. Tong, H., 1977, More on autoregressive model fitting with noisy data by Akaike's information criterion, IEEE Trans. Inform. Theory, IT-23: 409-410.
17. van den Bos, A., 1971, Alternative interpretation of maximum entropy spectral analysis, IEEE Trans. Inform. Theory, IT-17: 493-494.